基于增强学习的网格化出租车调度方法 *

何胜学

(上海理工大学 管理学院，上海 200093)

摘 要：高度信息化的网格化城市管理可以为出租车运营优化提供新的实时动态乘客需求信息和车辆位置信息。以此为契机，针对城市出租车空驶率高和司乘匹配率低的问题，提出了一种网格化的出租车实时动态调度的增强学习控制方法。通过为出租车提供空驶巡游的动态最佳路线，新的控制方法旨在提高出租车的服务效率，并降低乘客的等待时间。首先，以城市单元网格为基础，明确出租车调度的关键问题；其次，以空驶路线的动态调整为控制手段，建立调度的增强学习模型；最后，给出求解模型的 Q 学习算法，并通过算例验证新调度方法的有效性。研究表明新方法可以有效提高司乘匹配率、增加总的出租车运营收入、减少乘客平均等车时间和减少总的出租车空驶时间。

关键词：城市交通；出租车调度；增强学习；网格化管理；自适应式控制

中图分类号：U491 doi: 10.3969/j.issn.1001-3695.2017.11.0995

Grid-based taxi dispatching method based on reinforcement learning

He Shengxue

(*Business School*, *University of Shanghai for Science & Technology*, *Shanghai* 200093, *China*)

Abstract: Highly-informed grid-based city management can supply the real time passenger information and the position information of taxis for taxi operation optimization. On this account, we proposed a grid-based taxi dispatching dynamic control method based on reinforcement learning to solve the problem of the high vacant taxis rate and the low matching rate between taxis and passengers. By providing the optimal cruising routes for the vacant taxis, the new control method aims to improve the service level of taxis and to lower the waiting time of passengers. Firstly, based on the grids of city, we clarified the key problem of taxi dispatching. Secondly, by using the adjustment of vacant taxi route as the control action, we formulated the reinforcement learning model of taxi dispatching. At last, we proposed the corresponding Q learning algorithm to solve the new model. Numerical example demonstrated the effectiveness of the new dispatching method. The results show that the new method can not only increase the match rate between taxis and passengers and the total income of operation of taxi service, but also reduce the average waiting time of passengers and the total travel time of vacant taxis.

Key Words: urban transportation; taxi dispatching; reinforcement learning; grid management; adaptive control

0 引言

信息时代的迅猛发展为交通科学研究提供了新的契机与挑战。大数据技术的广泛应用使得过去对出租车出行需求的经验性估计演变为实时动态网络化分布的特征分析与预测。GPS 和 GIS 的发展为管理部门提供了对出租车的实时位置和行驶轨迹的更加全面信息。过去的扬招式出租车服务也逐渐被各种信息平台的 APP 服务所替代。上述变化为出租车公司或相关管理部门提供了进一步优化系统运作的机遇。如何有效利用实时车辆位置信息和乘客的出行需求分布实现出租车的实时调度优化业已成为出租车企业面对的首要技术问题。以网格化城市管理为依托，本文提出基于网格化出行需求信息更新和出租车路线优化的动态出租车调度控制方法。

下面简单介绍出租车调度的相关研究现状。合理的出租车调度可以有效减少出租车空驶产生的费用，有时减少的比例高达 90%[1,2]。早期研究重点关注出租车的合理定价和总体规模，而当前研究则涉及出租车服务系统的各个方面。其中主要包括网络规模的出租车运行建模[3~5]、随机的乘客出行需求 [6]、出租车电招系统[7,8]、基于元胞网络的出租车运行[9]、需求响应式出租服务[10]以及司乘匹配过程分析[8,11,12]。上述研究的共同之处均假设空载出租车的运行满足网络均衡、出行需求信息事先已知，并且出租车巡游速度给定。在均衡条件下，每一辆空载的出租车选择最近的可获得最大收益的区域作为行驶目的地，没有出租车能通过单方面改变空载行驶路线获得更高收益。上述

基金项目：上海理工大学人文社科攀登重点项目（SK17PA02）；上海市一流学科建设项目（S1201YLXK）

作者简介：何胜学（1976-），男，陕西三原人，副教授，博士，主要研究方向为交通网络建模（lovellhe@126.com）.

研究较少考虑动态变化的出行需求和路网实时交通状态的影响，因此很难满足当前高度信息化出租车市场发展的需要。

随着出租车叫车 APP 的大量涌现，出租车调度在提高司乘匹配率方面变得越来越重要[13]。针对出租车调度优化问题，研究者分别从车辆路径问题(vehicle routing problem,VRP)[14~16]和两分图匹配问题（bipartite graph matching problem,BGMP）[2,17,18]角度出发进行建模研究。一般而言，基于 VRP 的研究为每一辆出租车分配一系列的乘客；而基于 BGMP 的研究遵循就近原则匹配出租车和乘客。上述研究共同的不足是对出租车运行时间的随机性和乘客出行需求的随机性的考虑不足，因此实用性不强。

针对上述的现有研究不足，本文从四方面进行了改进：a）通过定义城市的网格化区块图，依据网格的乘客出行大数据(可从叫车 APP 平台和实际调查得到)建立各网格的出行率动态时间分布和目的地选择率。通过上述数据在控制方法实施中预测需求，决策空驶出租车路线。实际的出行需求也可通过实际的系统状态变量体现，并为调度方法所利用；b）为了更好地体现实时交通路网的交通状态对出租车巡游速度的影响，新控制方法可以在各个控制时刻利用实时路况信息和当前的出租车 GPS 位置更新达到目的地的时间(可通过重新计算最短路径实现)，从而实现对出租车巡游速度随机特征的把控，提升现有调度系统的可靠性；c）新的调度方法以出租车系统运行的仿真模型为基础，使得系统的状态演变更加细致，控制行为的选择更加准确。例如，对每个乘客旅行过程的各个关键时间点的精确描述；d）以增强学习理论为基础，新的控制方法可以有效地处理线上线下的优化学习，并可以根据系统的各种外在变化，如突增的出行需求，适应性地作出调整。

1　基本参变量和调度目标

1.1　基本参变量

T 表示离散时间步长,即实施调度控制的时间间隔。

k 表示时间分段标记，$k=0,1,\ldots,K$。$t=kT$ 表示实施调度控制的一个时间点，K 是实施调度控制的时间范围。

P 表示所有乘客集合, $p\in P$ 表示一个典型的乘客。

V 表示所有出租车集合, $v\in V$ 表示一辆典型的出租车。

G 表示所有单元网格的集合, $g\in G$ 表示一个典型的网格。

γ_g 表示网格 g 在一个时间分段内的乘客生成率。

$r_{g,h}$ 表示在网格 g 内生成的一个乘客选择另一网格 h 作为其目的地的概率。

g_p^O 表示乘客 p 的出发网格,即起点。

g_p^D 表示乘客 p 的目的地网格,即终点。

$t_{p,1}$ 表示乘客 p 的生成时刻，即该乘客利用 APP 请求服务的时间。

$t_{p,2}$ 表示乘客 p 在其出发网格的登车时刻。

$t_{p,3}$ 表示乘客 p 到达目的地时的下车时刻。

t_p^W 表示乘客 p 的等车时间。当有出租车满足其需求时，t_p^W 等于 $t_{p,2}-t_{p,1}$。

t_p^R 表示乘客 p 的坐车时间。当有出租车满足其需求时，t_p^R 等于 $t_{p,3}-t_{p,2}$。

d_p 表示乘客 p 的起点与终点之间的乘车距离。

η_v 表示出租车 v 的状态。当该车有乘客乘坐时，η_v 取值为 1；否则，为 0。

g_v 表示出租车 v 的标记网格。当该车有乘客 p 乘坐时，g_v 取值为 g_p^D；否则，为出租车 v 的实际所在网格。

t_v 表示出租车 v 的等待激活时间，即距离该车下次被实施调度的时间间隔长度。当该车有乘客 p 乘坐时，t_v 表示距离网格 g_p^D 的行程时间；否则，t_v 的值为 0。

$G_{g,n}$ 表示从网格 g 出发，在 n 个时间步长 T 内，一辆出租车可以达到的网格集合，即网格 g 的 n 级邻域。

$P_g(k)$ 表示网格 g 内在时刻 kT 等待出租车的乘客集合，$n_{g,P}(k)$ 表示集合 $P_g(k)$ 中的乘客总数；

$V_g^1(k)$ 表示时刻 kT 在网格 g 内空驶的出租车集合，$n_{g,V,1}(k)$ 表示 $V_g^1(k)$ 中的出租车总数。

$V_g^2(k)$ 表示时刻 kT 以网格 g 为目的地载客行驶的出租车集合，$n_{g,V,2}(k)$ 表示 $V_g^2(k)$ 中的出租车总数。

c_0 表示出租车的起步价格。

$\bar{c}$ 表示超出起步距离 d_0 后,出租车每公里的单价。

$\hat{t}$ 表示乘客在起点的最长可接受等车时间。

1.2　网格化出租车调度思想

图 1 给出了一个城市网格化的示意图。乘客 p 从其起点 g_p^O 出发沿虚线路径乘坐出租车达到目的地 g_p^D。任意乘客的乘车路线将由调度系统通过最短路径搜索算法得到。网格 g 的 1 级邻域有两种选择方式。一种是由与其紧邻的上下左右 4 个网格与其自身构成；另一种则进一步包括与其 4 个角相接的 4 个对角网格。网格的 2 级邻域则由其 1 级邻域包含的所有网格的 1 级邻域构成。第一种选择方式得到的邻域称为基础型邻域；第二种选择方式得到的邻域称为扩展型邻域。在算例分析时，本文将比较两种邻域构成方法对结果的影响。

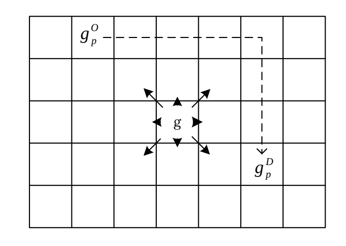


图 1　城市网格化图

需要区分两种乘客生成方式。实际应用时，当前乘客的信息完全由相应 APP 平台提供；而未来的乘客则由历史数据预测得到。在算例分析部分，本文通过给定的网格乘客生成率 γ_g 和目的地分布概率 $r_{g,h}$ 来仿真实现当前乘客信息和对未来乘客信息的预测。具体操作可分三步。首先依据 γ_g 随机生成当前时间阶段的乘客；然后根据 $r_{g,h}$ 利用轮盘赌规则确定每个新乘客的

目的地；最后利用最短路搜索算法确定乘客的乘车路线。

当一辆空驶出租车 v 与同一网格内的乘客 p 匹配后，出租车将按照乘客的乘车路线行驶。出租车的状态 η_v、g_v 和 t_v 随之更新。在任意给定的时间 kT，同一网格内的等待乘客与未载客出租车的匹配遵循如下两原则。乘客按照等待时间长短排序，排除等待时间超过 $\hat{t}$ 的乘客。排序后的乘客等待时间越长越早得到服务。假设所有出租车同质，在匹配中无优先权问题。匹配完成后更新各个网格的特征 $P_g(k)$、$V_g^1(k)$ 和 $V_g^2(k)$。当 $d_p \geq d_0$ 时乘客 p 的付费为 $c_p = c_0 + \bar{c}(d_p - d_0)$；否则，乘客 p 的付费为 c_0。因为假定出租车载客的行驶时间 t_v 会受到实际路网交通状态的影响，因此实际应用时 t_v 应当根据 GPS 提供的车辆位置和交通状态不断调整。在每一控制时间点 kT，根据载客车辆运行过程中 GPS 提供的车辆位置和网络中交通状况，以出租车的当前位置为起点，以所载乘客的目的地为终点，计算车辆的最佳行驶路线，并更新车辆预期抵达目的地的时间，即 t_v。在随后的算例分析中，为了模拟上述动态调整过程，本文假设 t_v 为随机变量。根据出租车 v 所载乘客 p 的 d_p 大小对 t_v 的大小做随机性处理。具体做法为设定一个单位距离的行驶时间随机误差量 χ 服从正态分布 $n(0,\ell)$，且车辆行驶各路段的行程时间相互独立。那么距离 d_p 的随机时间误差 χ_p 满足正态分布 $n(0,d_p\ell)$。假如车辆行驶距离 d_p 的期望时间为 $\bar{t}_v$，那么 $t_v = \bar{t}_v + \chi_p$。在仿真模拟时，通过为 χ_p 取满足其概率分布的一个随机量，具体确定 t_v 的值。

在乘客生成和司乘匹配过程中系统的运行基本是确定性的，而当可行的匹配完成后如何确定空载出租车的巡游路线则成为调度系统的最大挑战。在过去的扬招式出租车服务中，司机凭个人经验确定空驶巡游路线。但是这种方式会带有很大的随意性，往往造成部分区域出租车数量过多，而部分区域的乘客却得不到及时服务的情况。在信息高度发达的今天，出租车服务的 APP 化为本文提供了重新审视该问题，即从系统整体角度考虑优化出租车的空驶巡游路线的机会。下节将从增强学习角度详细分析该问题。

出租车调度的核心是确定空载车辆的巡游路线，而目的则是减少乘客的平均等车时间、提高司乘匹配率、减少车辆总的空驶时间和增加出租车运营收入。本研究以提高司乘匹配率为控制行为的直接价值评估指标。在数值算例分析中将对上述乘客的平均等车时间、司乘匹配率、车辆总的空驶时间和出租车运行总收入进行分析。

2 增强型学习模型

出租车调度系统的增强学习(RL)控制模型包括五个主要组成部分，即控制行为、控制信息、状态变量、状态价值函数和控制策略。

假设当前时刻为 kT。当每一个网格可行的司乘匹配完成后，可能会出现部分区域的乘客需要继续等待，而部分网格出现空载的出租车。此时，需要对空载出租车的巡游路线进行决策。设当前空载出租车 v 所在网格 g 的 1 级邻域为 $G_{g,1}$。那么 v 可以选择 $G_{g,1}$ 中的任意网格作为 $(k+1)T$ 时刻的目的地。令 $V(k) := \{v \mid v \in V, \eta_v(k) = 0\}$。对于 $v \in V(k)$，其选择的下一控制时刻 $(k+1)T$ 的目的地网格设为 $h_v(k)$。那么由所有 $h_v(k)$，$\forall v \in V(k)$ 构成的网格向量就是对应当前时刻 kT 的控制行为，表示为 $a(k)$。显然当集合 $V(k)$ 包含的元素较多时，$a(k)$ 的可行域 $A(k)$ 将非常庞大。基于增强学习的调度控制目的就是在每个控制时刻选取合理的控制行为，从而实现系统时空上的整体优化。

系统运行中伴随的随机或不确定因素被称为增强学习的控制信息。出租车调度中主要存在两种不确定性因素。一种是单元网格中乘客生成的随机性；另一种是出租车载客行程时间的不确定性。前者在模型中由乘客生成率 γ_g 和目的地的分布概率 $r_{g,h}$ 确定，而后者由具有随机特征的出租车 v 的等待激活时间 t_v 来体现。但为了表述方便，将上述随机因素用抽象向量 Δ 表示。向量 Δ 就是系统所需的控制信息。

在每一个控制时间点，控制系统需要基于系统当前的状态和相关的控制信息确定最佳的控制行为。系统的状态变量就是一组描述系统给定时刻状态的特征向量。全面细致地描述出租车服务系统需要对系统大量的细节加以刻画。但是这不仅会造成状态变量过于繁琐，也会使随后的控制行为决策为繁复的次要因素所困。因此本研究将以网格为对象，主要考虑三个特征量，即 $n_{g,P}(k)$、$n_{g,V,1}(k)$ 和 $n_{g,V,2}(k)$。

假设每次控制行为决策前，每个网格内所有可行的司乘匹配均已完成。此时对于任意网格 g，其对应的等待乘客数目 $n_{g,P}(k)$ 和空载出租车数目 $n_{g,V,1}(k)$ 不能同时为正。本文选择司乘匹配完成后的 $n_{g,P}(k)$、$n_{g,V,1}(k)$ 和 $n_{g,V,2}(k)$ 作为 kT 时刻网格 g 状态特征。将所有网格在 kT 时刻的状态特征整合为系统状态向量

$$s(k) := \{\cdots, n_{g,P}(k), n_{g,V,1}(k), n_{g,V,2}(k), \cdots\} \tag{1}$$

$s(k)$ 就是对应控制阶段 k 的系统状态变量。所有可行系统状态变量值构成系统状态空间集合 S。

控制策略指的是由系统状态变量决定控制行为的一种决策函数，即任意给定一个系统状态，依据控制策略可得到对应的唯一控制行为。将任意一个控制策略表示为 $\pi \in \Pi$，其中 Π 为所有可行策略集合。控制策略的函数形式可表示为

$$a = A^{\pi}(s) \tag{2}$$

增强学习模型的目的就是在给定约束条件下确定可实现一定目的的最佳策略。本研究的目的之一就是确定实现提高司乘匹配率的出租车空载巡游路线与网格状态量之间的合理关联关系。

对于任意控制阶段，控制行为一旦确定必然会对系统当前和后续的运行产生影响。一般而言控制行为的当前影响较易度量，比如本文中空载车辆下一时刻的目的地一旦确定，随之增加的司乘匹配数就会有所变化。由于系统运行环境不断变化，控制行为的远期系统影响往往难以准确估计。在增强学习中，

本文将控制行为直接导致的系统优化目标的变化量称为控制行为的收益，对应的行为收益函数表示为 $f(s(k),a(k),\Delta(k+1))$。

行为收益函数 $f(s(k),a(k),\Delta(k+1))$ 中的 $\Delta(k+1)$ 表示从阶段 k 到阶段 $k+1$ 系统演变过程中可变为已知的随机控制信息。计算收益函数值的过程分四步。首先在当前阶段完成各个网格的所有可行司乘匹配，确定一个具体控制行为 $a(k)$；其次，在已实现的 $\Delta(k+1)$ 基础上，更新阶段 $k+1$ 各个网格的空载出租车数；接着，在各个网格实现所有可能的等车乘客与刚刚变为空载的车辆的匹配；最后在各网格实现剩余等车乘客与空载车辆的司乘匹配，记录该类匹配的总数，即对应控制行为 $a(k)$ 的收益函数值。上述计算的依据为刚完成载客任务的车辆就处于对应的网格，而其他的空载车辆则是由上阶段空载车辆经过路线控制(即控制行为 $a(k)$ 的作用)后得到的。

合理的控制决策必须考虑控制行为对系统后续状态的影响，以及这种影响的时间衰减性。因此选取如下的优化目标：

$$\max_{\pi\in\Pi}\mathbb{E}\{\sum_{k=0}^{\infty}[\gamma^k f(s(k),A^{\pi}(s(k)),\Delta(k+1))]\} \tag{3}$$

其中：$\gamma<1$ 为折扣因子，算子 $\mathbb{E}$ 表示求变量的期望值。直接求解优化问题(3)非常困难，因此通常的做法是转而求解该问题等价的 Bellman 方程。为了简化公式表述，用下标"k"取代变量的时间标签"(k)"。设系统处于状态 s 时的状态价值为 $V(s)$。依据 Bellman 方程，最优状态价值 $V^*(s)$ 应满足：

$$V^*(s_k)=\mathbb{E}[-f(s_k,A^*(s_k),\Delta_{k+1})+\gamma V^*(s_{k+1}\,|\,s_k,A^*(s_k),\Delta_{k+1})] \tag{4}$$

下文将设计对应的 Q 学习算法求解上述问题。$V(s)$ 与 Q 函数 $Q(s,a)$ 具有如下关系：

$$V(s)=\max_a Q(s,a)\,. \tag{5}$$

与 $V(s)$ 相比，$Q(s,a)$ 的变量维度增加了一位；但是实际应用时可通过比较不同控制行为的 $Q(s,a)$ 值更加方便地确定最佳行为。问题的求解依赖于 $Q(s,a)$ 的具体形式，但是目前不存在 $Q(s,a)$ 的任何具体形式。因此需要利用某种带有待定参数 θ 的近似函数 $Q(s,a,\theta)$ 来替代 $Q(s,a)$。最常见的做法是利用人工神经网络(ANN)近似技术完成上述任务。此时参数 θ 表示神经网络的联接权重和节点阈值[19,20]。根据 ANN 基本理论，利用多层的 ANN 即可实现对各种函数的无限逼近。特别是当具体函数形式未知条件下，ANN 的规范统一结构和有效地参数调整机制为研究提供了一种便捷有效函数近似工具。

3 Q 学习算法

Q 学习算法是求解增强学习模型的一种非常有效算法[19]。通过将状态价值函数转换为 Q 函数，可以在算法执行过程中通过直接比较 Q 函数值的大小选取最佳的控制行为。通过迭代学习，当 Q 函数的近似形式接近实际 Q 函数时，上述的控制行为选取变得更加有效。而利用状态价值函数确定最佳控制行为时，必须针对每一个可选行为，将系统进行状态转移，从而确定下一状态的价值函数值。通过比较这些可能的后续状态价值函数值，确定最佳控制行为。与 Q 学习算法相比，利用状态价值函数来确定控制行为不仅繁琐，而且系统状态转移所带来的计算量也非常可观。因此，本文选取 Q 学习算法作为求解增强学习模型的基本算法。

在算法中 k 代表当前阶段的序列号，显然 kT 对应一个控制时刻。m 表示对系统进行仿真模拟的序号。K 和 M 分别是总的阶段数目和模拟仿真总次数。针对网格化出租车调度优化的 Q 算法的求解具体步骤如下：

a)初始化。对所有的系统状态 s_k，给出价值函数 $\bar{Q}_k^0(s_k,a_k)$ 的近似值以及控制行为 $a_k\in A(k)$，$k=\{0,1,\ldots,K-1\}$。令 $m=1$，并初始化 $s^1(k)$。

b)选择一个随机信息的样本路径 ω^m。

c)将下面操作步骤依 $k=0,1,\ldots,K-1$ 加以循环迭代:

(a)利用 ε 贪婪规则确定控制行为。以概率 ε，从控制行为集 $A(k)$ 中随机选择行为 a_k^m。而以概率 1-ε，利用公式 $a_k^m\in\arg\max_{a_k\in A(k)}\bar{Q}_k^{m-1}(s_k^m,a_k,\theta)$ 选择行为 a_k^m。

(b)对信息 $\Delta_{k+1}^m:=\Delta(\omega_{k+1}^m)$ 进行取样，并计算状态变量 $s_{k+1}^m=S^M(s_k^m,a_k^m,\Delta_{k+1}^m)$。

(c)计算 $\hat{q}_k^m=-f(s_k^m,a_k^m,\Delta_{k+1})+\gamma\max_{a_{k+1}\in A(k+1)}\bar{Q}_{k+1}^{m-1}(s_{k+1}^m,a_{k+1},\theta)$。

(d)更新参数 $\theta:=\theta+\lambda\nabla\bar{Q}_k^{m-1}(-f(s_k^m,a_k^m,\Delta_{k+1})+\gamma\bar{Q}_{k+1}^{m-1}-\bar{Q}_k^{m-1})$。其中的控制行为 a_{k+1} 可由 ε 贪婪规则确定。其次，基于更新后的参数 θ，按照下式计算 Q 因子：

$$\bar{Q}_k^m(s_k^m,a_k^m)=(1-\alpha_{m-1})\bar{Q}_k^{m-1}(s_k^m,a_k^m,\theta)+\alpha_{m-1}\hat{q}_k^m\,。$$

d)令 $m:=m+1$。如果 $m\le M$，转步骤 b)。

e)给出最终的 Q 因子 $(\bar{Q}_k^M)_{k=1}^{K-1}$ 和 θ。

上述算法的目的可看做最小化 θ 的函数 $z(\theta)=(1/2)(\hat{q}_k^m-\bar{Q}_k^{m-1}(s_k^m,a_k^m,\theta))^2$。函数 $z(\theta)$ 的负梯度方向为 $\nabla\bar{Q}_k^{m-1}(s_k^m,a_k^m,\theta)(\hat{q}_k^m-\bar{Q}_k^{m-1}(s_k^m,a_k^m,\theta))$

其中：θ 作为自变量，而 $\hat{q}_k^m$ 为给定值。按照经典非线性规划理论，在自变量 θ 的可行域内，如果当前的 θ 对应的函数值非局部或全局最小值，且步长足够小，目标函数 $z(\theta)$ 的值将沿当前变量 θ 处的负梯度方向下降。因此算法中的参数 λ 相当于沿可行下降方向的搜索步长。可以通过试错法确定 λ 的合理取值。

4 数值实验

本章以图 2 中由 15 个网格构成的出租车调度区域为例对本文给出方法进行验证分析。15 个网格的编号和对应的乘客生成率 γ_g 已在图中标出。网格的长和宽均为 1000 m。离散时间步长 T 为 100 s，总的时间分段数 K 设为 100。出租车的平均速度为 36 km/hr(即 10 m/s)。出租车的起步费 c_0 为 14 元，起步距离

d_0 为 3 km，而超出起步距离后每公里的单价 $\bar{c}$ 为 2.5 元。乘客在起点的最长可接受等车时间 $\hat{t}$ 设为 400s。假设网络中运行的出租车总数为 30 辆。总的仿真次数 M 为 300 次，参数 ε 和折扣因子 γ 均设为 0.5，算法的步长 λ 设为 0.01。ε 值的大小决定了学习过程中的利用已有经验进行控制行为选取和控制行为随机选取的占比。ε 值的较大，控制行为随机选取的几率就大，算法运行的前期可能需要更多的迭代得到较好的系统表现，但是后期的最佳控制行为选择会更有效。而当 ε 值的较小时，在算法运行的前期，系统表现较好，但需要更多的迭代实现后期的系统表现提升[19]。γ 作为目标的时间折扣因子，其大小反映了算法对系统表现随时间变化的一种折现评价[19]。上述 ε 和 γ 值的选择只是各种可能性的一种，实际应用时应通过试算确定其合理取值。上述 λ 步长的选择也只是一个示例，实际算法应用时，应通过试算确定合理取值。一般而言当 Q 函数值较大时，λ 的取值应当较小。否则，由于目标函数的剧烈变化，算法在执行初期将极不稳定[20]。假设出租车行驶 1 公里可能产生的行驶时间误差服从均值为 0 而标准方差为 20 s 的正态分布。

1 (0.2)	2 (0.3)	3 (0.6)	4 (0.2)	5 (0.5)
6 (0.2)	7 (0.9)	8 (0.4)	9 (0.6)	10(0.4)
11(0.2)	12(0.3)	13(0.6)	14(0.2)	15(0.3)

图 2　具有 15 个网格的调度区域

为了使随后的图表更加清晰简洁，定义如下的符号变量。n_{TC} 表示总的司乘匹配数；n_{NS} 表示未得到服务而损失的乘客数；$\bar{t}_W$ 表示乘客平均等车时间(s)；I 表示总的出租车运营收入(元)；T_{NO} 表示总的出租车空驶时间(s)。NC 表示无控制情景下的出租车运营；CBN 表示基于基础型邻域的调度控制；CEN 表示基于扩展型邻域的调度控制。

表 1　不同控制情景下的优化结果比较

控制方法	n_{TC}	n_{NS}	$\bar{t}_W$	I	T_{NO}
NC	719	235	135.3	10991	69400
CBN	845	139	130.3	12725	47200
CEN	983	53	105.9	14393	54200

表 1 给出了三种不同控制方式下，出租车服务系统的运行表现。控制情景下的数据来自系统在 300 次仿真后，排除 ε 贪婪机制而仅采用最佳控制行为的运算结果。可以看出，与无控制情景相比，调度控制可以明显改善系统的运行效率；而基于扩展型邻域的调度在多个方面优于基于基础型邻域的调度。因为优化的直接目标是增加司乘的匹配数，因此基于扩展型邻域的调度需要出租车空驶更多的时间实现更多的司乘匹配。这可以解释为什么表 1 中 CEN 的总出租车空驶时间 T_{NO} 比 CBN 要高。

下面以基础型邻域控制 CBN 为例，分析随着仿真次数的增加各个系统运行指标的变化情况。

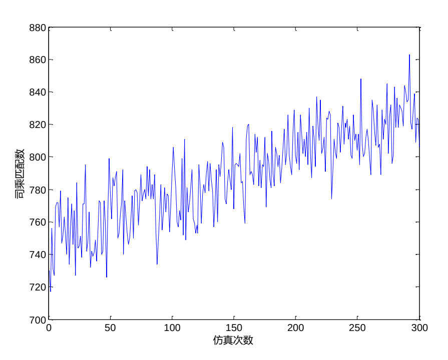


图 3　司乘匹配数的变化情况

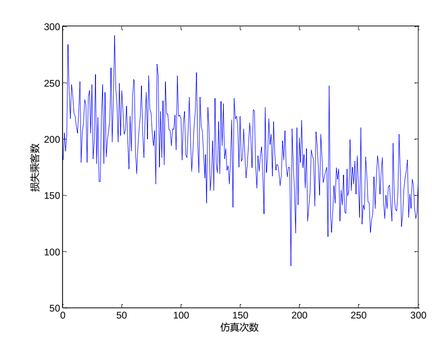


图 4　未得到服务而损失的乘客数的变化情况

图 3 和 4 分别给出了司乘匹配数 n_{TC} 和损失的乘客数 n_{NS} 随仿真次数增加的变化情况。随着仿真次数的增加，可以看出司乘匹配数 n_{TC} 呈现出上升的趋势，而损失的乘客数 n_{NS} 呈现下降趋势。而具体数值的随机波动变化源于系统本身的随机特征和算法中 ε 贪婪机制选择控制行为的随机性。

图 5 中乘客平均等车时间 $\bar{t}_W$ 随着仿真次数的增加在 130 秒上下随机波动。这说明仿真次数的增加并不能明显改善乘客的平均等车时间。但是本文发现利用扩展型邻域可以降低平均等车时间，而控制条件下平均等车时间均低于无控制的情景。

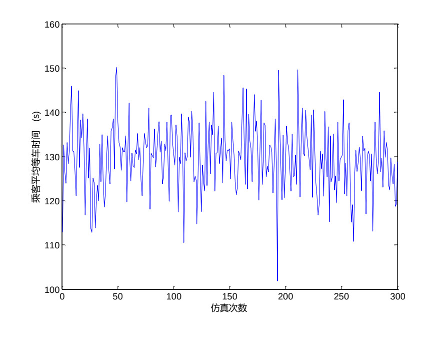


图 5　乘客平均等车时间 $\bar{t}_W$ 的变化(时间单位:秒)

图 6 中的总运营收入随着仿真次数的增加呈现出明显的增加趋势。图 7 中总的出租车空驶时间随着仿真次数的增加呈现出明显的减少趋势。这些变化与图 3 中司乘匹配数的增加趋势相对应。

算例求解的计算机程序用 Java 1.8.0 编写，在 NetBeans IDE 8.0.2 开发环境下实现，所用计算机处理器为 Intel® Core i3-3120M CPU。两种控制方式下完成 300 次系统仿真的计算时间

均为 14s，而完成一次仿真的计算时间约为 0.04s。

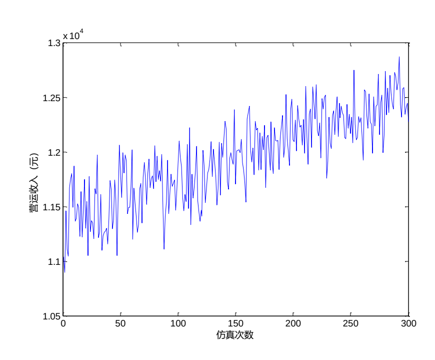


图 6　总的出租车运营收入 I 的变化(单位:元)

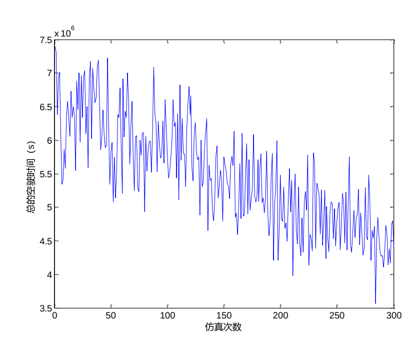


图 7　总的出租车空驶时间 T_{NO} 的变化（单位:s)

5　结束语

在网格化城市管理的背景下，针对网格化的出租车出行需求动态数据和网格化的出租车路线规划，提出了一种网络出租车调度的增强学习控制方法。新的控制方法可以有效处理系统的随机动态特征(包括随机的行程时间和出行需求)，通过无监督的自适应式强化学习实现出租车的空车路线调度。通过定义网格和网格邻域概念，使得在实施控制方法时可有效利用基于网格大数据的出租车出行需求特征与需求预测。动态的路线调整过程也使得利用实时的车辆定位信息成为现实。在出租车服务系统运行表现上，新的调度控制方法不仅可以增加司乘匹配数和降低乘客流失风险，而且可以增加出租车总的运营收入和降低乘客平均等车时间。

本文研究可从多个方面加以拓展，包括考虑出租车的不同类别、乘客的优先级别、网格划分的不同方式以及实际道路交通状态的实时影响分析等。同时研究方法的有效性还有待进一步的实证分析改进。

参考文献：

[1] 祝进城，帅斌，孙朝苑，等. 固定费率下城市出租车拥挤收费模型与算法 [J]. 计算机应用研究, 2013, 30 (8): 2288-2291.

[2] Zhan X, Qian X, Ukkusuri S V. A graph-based approach to measuring the efficiency of an urban taxi service system [J]. IEEE Trans on Intelligent Transportation Systems, 2016, 17 (9): 2479-2489

[3] 胡继华，黄泽，邓俊. 考虑出行方向的广州出租车时空可达性研究 [J]. 计算机应用研究, 2014, 31 (2): 454-456.

[4] Jung J, Chow J Y, Jayakrishnan R, et al. Stochastic dynamic itinerary interception refueling location problem with queue delay for electric taxi charging stations [J]. Transportation Research Part C, 2014, 40 (1): 123–142.

[5] Sayarshad H R, Chow J Y. Survey and empirical evaluation of non-homogeneous arrival process models with taxi data [J]. Journal of Advanced Transportation, 2016, 50 (7): 1275–1294.

[6] Zhang W, Ukkusuri S V. Optimal fleet size and fare setting in emerging taxi markets with stochastic demand [J]. Computer-Aided Civil and Infrastructure Engineering, 2016, 31 (9): 647-660.

[7] He F, Shen Z M. Modeling taxi services with smart phone-based e-hailing applications [J]. Transportation Research Part C, 2015, 58 (1): 93–106.

[8] Wang X, He F, Yang H, et al. Pricing strategies for a taxi-hailing platform [J]. Transportation Research Part E, 2016, 93 (2): 212–231.

[9] Wong R, Szeto W, Wong S. A cell-based logit-opportunity taxi customer-search model [J]. Transportation Research Part C, 2014, 48 (1): 84–96.

[10] Amirgholy M, Gonzales E J. Demand responsive transit systems with time-dependent demand: user equilibrium, system optimum, and management strategy [J]. Transportation Research Part B, 2016, 92 (2): 234–252.

[11] Yang T, Yang H, Wong S C. Taxi services with search frictions and congestion externalities [J]. Journal of Advanced Transportation, 2014, 48 (6): 575–587.

[12] Zha L, Yin Y, Yang H. Economic analysis of ride-sourcing markets [J]. Transportation Research Part C, 2016, 71 (2): 249–266.

[13] Nie Y M. How can the taxi industry survive the tide of ride sourcing?Evidence from Shen Zhen, China [J]. Transportation Research Part C, 2017, 79 (2): 242–256.

[14] Pillac V, Gendreau M, Gueret C, et al. A review of dynamic vehicle routing problems [J]. European Journal of Operations Research, 2013, 225 (1): 1–11.

[15] Hosni H, Naoum-Sawaya J, Artail H. The shared-taxi problem: formulation and solution methods [J]. Transportation Research Part B, 2014, 70 (3): 303–318.

[16] Jung J, Jayakrishnan R, Park J Y. Dynamic shared-taxi dispatch algorithm with hybrid-simulated annealing [J]. Computer-Aided Civil and Infrastructure Engineering, 2016, 31 (4): 275–291.

[17] Agatz N, Erera A, Savelsbergh M, et al. Optimization for dynamic ride-sharing: a review [J]. European Journal of Operations Research, 2012, 223 (2): 295–303.

[18] Nourinejad M, Roorda M J. Agent based model for dynamic ride sharing [J]. Transportation Research Part C, 2016, 64 (1): 117–132.

[19] Warren B P. Approximate dynamic programming-solving the curses of dimensionality [M]. Hoboken: Wiley, 2011.

[20] Dimitri P B, John N T, John T. Neuro-dynamic programming [M]. Nashua: Athena Scientific, 1996.